A Logical Theory of Belief Dynamics for Resource-Bounded Agents

Philippe Balbiani
IRIT-CNRS
Toulouse University, France
Philippe.Balbiani@irit.fr

David Fernández-Duque
IRIT, Toulouse University
France
David.Fernandez@irit.fr

Emiliano Lorini
IRIT-CNRS
Toulouse University, France
Emiliano.Lorini@irit.fr

ABSTRACT

The paper presents a new logic for reasoning about the formation of beliefs through perception or through inference in non-omniscient resource-bounded agents. The logic distinguishes the concept of explicit belief from the concept of background knowledge. This distinction is reflected in its formal semantics and axiomatics: (i) we use a non- standard semantics putting together a neighbourhood semantics for explicit beliefs and relational semantics for background knowledge, and (ii) we have specific axioms in the logic highlighting the relationship between the two concepts. Mental operations of perceptive type and inferential type, having effects on epistemic states of agents, are primitives in the object language of the logic. At the semantic level, they are modelled as special kinds of model-update operations, in the style of dynamic epistemic logic (DEL). Results about axiomatization, decidability and complexity for the logic are given in the paper.

Keywords

Epistemic logic; cognitive agents; resource-bounded reasoning

1. INTRODUCTION

Most of existing logical theories of epistemic attitudes developed in the area of epistemic logic assume that agents are omniscient, in the sense that: (i) their beliefs are closed under conjuction or under known implication, *i.e.*, if φ is believed and ψ is believed then $\varphi \wedge \psi$ is believed and if φ is believed and $\varphi \rightarrow \psi$ is believed then ψ is believed; (ii) their explicit beliefs are closed under logical consequence (*alias* valid implication), *i.e.*, if φ is believed and φ logically implies ψ, *i.e.*, $\varphi \rightarrow \psi$ is valid, then ψ is believed as well; (iii) they believe valid sentences or tautologies; (iv) they have introspection over their beliefs, *i.e.*, if φ is believed then it is believed that φ is believed.

As pointed out by [13, 16], relaxing the assumption of logical omniscience allows for a resource-bounded agent who might fail to draw any connection between φ and its logical consequence ψ and, consequently, to believe any valid sentence and who might need time to infer and form new beliefs from her existing knowledge and beliefs.

The aim of this paper is to propose a new logic which helps in clarifying how a non-omniscient resource-bounded agent can form new beliefs either through perception or through inference from her existing knowledge and beliefs. More precisely, the aim of the paper is to introduce a dynamic logic, called DLEK (Dynamic Logic of Explicit Beliefs and Knowledge) in which programs are mental operations, either of perceptive type or of inferential type, having effects on epistemic states of resourced-bounded agents.

This is not the first attempt to build a logic of epistemic attitudes for non-omniscient agents. Logics of awareness have been studied in the recent years (see, *e.g.*, [19, 12, 10, 1]) starting from the seminal work of Fagin & Halpern [7]. These logics distinguish between explicit beliefs that are under the focus of attention and implicit beliefs, namely, potential beliefs that are derivable from what an agent explictly believes. However, the crucial difference between DLEK and existing logics of awareness is that the former provides a constructive theory of explicit beliefs, as it accounts for the perceptive and inferential steps leading from an agent's knowledge and beliefs to new beliefs. A notable exception is [21], although our conceptual framework is different (see Section 2) and, unlike the present paper, the author does not provide any axiomatization or decidability result for his logic of reasoning steps, as he only provides a semantics. Moreover, [21] does not distinguish the concept of explicit belief and the concept of background knowledge, which is the fundamental distinction of our logic DLEK.

DLEK is the first logical theory of the relationship between explicit beliefs and background knowledge, both from a static and from a dynamic perspective. This is reflected in its formal semantics and axiomatics: (i) we use a non-standard semantics putting together a neighbourhood semantics for explicit beliefs and relational semantics for background knowledge, and (ii) we have specific axioms in the logic highlighting the relationship between the two concepts.

Our constructive approach to explicit beliefs also distin-

guishes DLEK from existing logics of time-bounded reasoning which represent reasoning as a process that requires time (see, e.g., [2, 9]). Specifically, existing logics of time-bounded reasoning account for the formation of new beliefs due to the time-consuming application of inference rules to the beliefs that are under the focus of attention. However, differently from DLEK, they do not include mental operations of perceptive type and inferential type as primitives in the object language of the logic. As we will show in the paper, the advantage of having mental operations in the object language of DLEK is that we can use it to reason about the consequences of a sequence of perceptive and inferential steps on the epistemic states of agents.

The paper is organized as follows. In Section 2 we present the conceptual foundation of our logic DLEK, namely the general view about dynamics of beliefs in resource-bounded agents which underlies our formal theory. The general idea is that new beliefs can be formed either by perception or by inferring them from existing beliefs in working memory and by retrieving information from background knowledge in long-term memory. Section 3 presents the syntax and the semantics of DLEK. We will show that in DLEK perceptive and inferential steps are modelled as special kinds of model-update operations, in the style of dynamic epistemic logic (DEL) [20]. In Section 4 we present an example illustrating the expressive power of the logic. Section 5 is devoted to present an axiomatics. Finally, Section 6 presents complexity results for the satisfiability problem of the static fragment of DLEK, called LEK as well as a decidability result for the satisfiability problem of DLEK.

2. CONCEPTUAL FRAMEWORK

The cognitive architecture underlying the logic DLEK (Dynamic Logic of Explicit Beliefs and Knowledge) is represented in Figure 1. It clarifies the processing of information in human agents and human-like artificial agents.

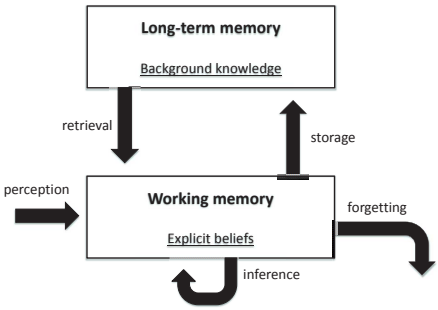

Figure 1: Cognitive view of the relationship between background knowledge and explicit beliefs

In accordance with existing psychological theories and computational models of memory and attention [3, 6, 22, 17], we assume that an agent has two kinds of information in her mind, those available in long-term memory (LTM) and those directly accessible in working memory (WM).

The information available in long-term memory, generally called *background knowledge*, includes both knowledge of specific events in the past and conceptual or causal knowledge representing the agent's unproblematic interpretation of reality.[1] For example, an agent may have background conceptual knowledge about how restaurants are organized or background causal knowledge about the relation between smoke and fire. In particular, she may know that restaurants have waiters, chairs and tables or that if smoke comes out from a window of a certain house, then there is fire inside the house.

Working memory retains information in an accessible state suitable for carrying out any kind of reasoning or decision task. In particular, following [14], we assume that the information available in an agent's working memory includes all *explicit beliefs* of the agent that occupy her consciousness and draw on her limited capacity of attention.[2] Some explicit beliefs are formed via *perception*. Formation of explicit beliefs via perception just consists of adding a new belief to the set of beliefs that are under the focus of the agent's attention. For example, an agent may look outside the window, see that it is raining outside, and thereby start believing that it is raining outside.

An agent can also use her explicit beliefs as premises of an *inference* which leads to the formation of a new belief. In some cases, formation of explicit beliefs via inference requires the *retrieval* of information from long-term memory. For example, suppose that an agent sees that smoke comes out from the window of a certain house and, as a result, she starts to explicitly believe this. The agent retrieves from her background knowledge stored in her long-term memory the information that if smoke comes out from a window of a certain house, then it means that there is fire inside the house. The agent can use this information together with the belief that smoke comes out from the window available in her working memory, to infer that there is fire inside the house and to form the corresponding belief.

Information can also be lost from working memory through *forgetting*: an agent may explicitly believe something but not believe it anymore at a later point. Information can also be removed from working memory and stored in long-term memory to make it available for a later moment. *Storage* of information in long-term memory might be necessary, given the limited capabilities of working memory.

In the next section we present the syntax and the semantics of the logic DLEK which makes precise all concepts informally discussed in this section.

3. LOGICAL FRAMEWORK

DLEK is a logic which consists of a static component and a dynamic one. The static component, called LEK, is a logic of explicit beliefs and background knowledge. The dynamic component extends the static one with dynamic operators capturing the consequences of the agents' mental operations on their explicit beliefs.

3.1 Syntax

Assume a countable set of atomic propositions $Atm =$

[1] In the Soar architecture [15] these two kinds of background are called, respectively, episodic memory and semantic memory.

[2] Some psychologists (e.g., [6]) distinguish focus of attention from working memory, as they assume that there might be information activated in working memory which is not under the focus of the agent's attention. For simplicity, we here assume that focus of attention and working memory are coextensive.

$\{p, q, \ldots\}$ and a finite set of agents $Agt = \{1, \ldots, n\}$. The set of groups (or coalitions) is defined to be $2^{Agt*} = 2^{Agt} \backslash \{\emptyset\}$. Elements of 2^{Agt*} are denoted by $J, J', \ldots$ We denote with *Prop* the set of all Boolean formulas built out of the set of atomic propositions *Atm*.

The language of DLEK, denoted by $\mathcal{L}_{\mathsf{DLEK}}$, is defined by the following grammar in Backus-Naur Form:

$$\begin{array}{lll} \alpha & ::= & \vdash(\varphi,\psi) \mid \cap(\varphi,\psi) \mid +\varphi \mid -\varphi \\ \varphi, \psi & ::= & p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{B}_i\varphi \mid \mathsf{K}_i\varphi \mid [\alpha]\varphi \end{array}$$

where p ranges over *Atm* and i ranges over *Agt*.

The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from p, $\neg$ and $\wedge$ in the standard way.

The language of LEK, the fragment of DLEK without dynamic operators, is denoted by $\mathcal{L}_{\mathsf{LEK}}$ and defined by the following grammar in Backus-Naur Form:

$$\varphi, \psi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{B}_i\varphi \mid \mathsf{K}_i\varphi$$

where p ranges over *Atm* and i ranges over *Agt*. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from p, $\neg$ and $\wedge$ in the standard way. In what follows, we explain the meaning of the operators of our logic.

The formula $\mathsf{B}_i\varphi$ is read "the agent i explicitly believes that φ is true" or, more shortly, "the agent i believes that φ is true". As explained in Section 2, explicit beliefs are accessible in working memory and are the basic elements of the agents' reasoning process.

The modal operator K_i captures the notion of background knowledge discussed in Section 2. It represents the information that agent i *can* use and that is available in long term memory to infer new explicit beliefs. Some pieces of background knowledge are either conceptual or causal and represent the agent's unproblematic interpretation of reality.

Differently from explicit beliefs, background knowledge is assumed to satisfy 'omniscience' principles like closure under conjuction and known implication, closure under logical consequence and introspection. Specifically, as we will show below, K_i is nothing but the well-known S5 operator for knowledge widely used in computer science [8]. The fact that the operator K_i satisfies 'omniscience' principles is justified by the assumption that the information that an agent possesses in background defines a deductively closed *knowledge base*.

The formula $[\alpha]\psi$ has to be read "ψ holds, after the mental operation (or mental action) α is publicly performed by all agents". We distinguish four types of mental operations α which allow us to capture some of the dynamic properties of explicit beliefs and background knowledge informally described in Section 2 above: $+\varphi$, $-\varphi$, $\vdash(\varphi,\psi)$ and $\cap(\varphi,\psi)$.

$+\varphi$ and $-\varphi$ correspond, respectively, to the mental operations of forming an explicit belief via perception and forgetting an explicit belief represented in Figure 1.

$\vdash(\varphi,\psi)$ and $\cap(\varphi,\psi)$ characterize two basic operations of forming explicit beliefs via inference. Specifically, $\vdash(\varphi,\psi)$ is the mental operation which consists in inferring ψ from φ in case φ is believed and, according to an agent's background knowledge, ψ is a logical consequence of φ. In other words, by performing this mental operation, an agent tries to retrieve from her background knowledge in long-term memory the information that φ implies ψ and, if she succeeds, she starts to believe ψ. $\cap(\varphi,\psi)$ is the mental operation which consists in closing the explicit belief that φ and the explicit belief that ψ under conjunction. In other words, $\cap(\varphi,\psi)$ characterizes the mental operation of deducing $\varphi \wedge \psi$ from the explicit belief that φ and the explicit belief that ψ.

In this paper we assume that, differently from explicit beliefs, background knowledge is irrevocable in the sense of being stable over time [5]. In the conclusion, we will offer some insights on how to make background knowledge dynamic by including in our logic the operation of storing information in long-term memory, represented in Figure 1.

3.2 Semantics

The main notion in semantics is given by the following definition of LEK model which provides the basic components for the interpretation of the static logic LEK:

DEFINITION 1 (LEK MODEL). *A LEK model is a tuple* $M = (W, N, R_1, \ldots, R_n, V)$ *where:*

- W *is a set of worlds or situations;*
- *for every* $i \in Agt$, $R_i \subseteq W \times W$ *is an equivalence relation on* W*;*
- $N : Agt \times W \longrightarrow 2^{2^W}$ *is a neighbourhood function such that for all* $i \in Agt$, $w, v \in W$ *and* $X \subseteq W$*:*

  (C1) *if* $X \in N(i, w)$ *then* $X \subseteq R_i(w)$,
  (C2) *if* wR_iv *then* $N(i, w) = N(i, v)$*;*
- $V : W \longrightarrow 2^{Atm}$ *is a valuation function.*

For every $i \in Agt$ and $w \in W$, $R_i(w) = \{v \in W : wR_iv\}$ identifies the set of situations that agent i considers possible at world w. In cognitive terms, $R_i(w)$ can be conceived as the set of all situations that agent i *can* retrieve from her long-term memory and reason about them. More generally, $R_i(w)$ is called agent i's epistemic state at w. The reason why R_i is an equivalence relation is that it is used to model a form of omniscient background knowledge instead of omniscient background belief. The latter could be modelled by replacing the equivalence relations R_i by serial, transitive and Euclidean relations commonly used in doxastic logic to model a notion of belief.

For every $i \in Agt$ and every $w \in W$, $N(i, w)$ defines the set of all facts that agent i explicitly believes at world w, a fact being identified with a set of worlds. More precisely, if $A \in N(i, w)$ then, at world w, agent i has the fact A under the focus of her attention and believes it. $N(i, w)$ is called agent i's explicit belief set at world w.

Constraint (**C1**) just means that an agent can have explicit in her mind only facts which are compatible with her current epistemic state. According to Constraint (**C2**), if world v is compatible with agent i's epistemic state at world w, then agent i should have the same explicit beliefs at w and v.

Truth conditions of DLEK formulas are inductively defined as follows.

DEFINITION 2 (TRUTH CONDITIONS). *Let* $M = (W, N, R_1, \ldots, R_n, V)$ *be a LEK model. Then:*

$$\begin{array}{rcl} M, w \models p & \Longleftrightarrow & p \in V(w) \\ M, w \models \neg\varphi & \Longleftrightarrow & M, w \not\models \varphi \\ M, w \models \varphi \wedge \psi & \Longleftrightarrow & M, w \models \varphi \text{ and } M, w \models \psi \\ M, w \models \mathsf{B}_i\varphi & \Longleftrightarrow & ||\varphi||^M_{i,w} \in N(i, w) \\ M, w \models \mathsf{K}_i\varphi & \Longleftrightarrow & M, v \models \varphi \text{ for all } v \in R_i(w) \\ M, w \models [\alpha]\varphi & \Longleftrightarrow & M^\alpha, w \models \varphi \end{array}$$

where

$$||\varphi||^M_{i,w} = \{v \in W : M, v \models \varphi\} \cap R_i(w)$$

and $M^\alpha = (W, N^\alpha, R_1, \ldots, R_n, V)$ *such that, for all* $i \in Agt$ *and* $w \in W$*:*

$$N^{+\psi}(i,w) = N(i,w) \cup \{||\psi||^M_{i,w}\}$$

$$N^{-\psi}(i,w) = N(i,w) \setminus \{||\psi||^M_{i,w}\}$$

$$N^{\vdash(\psi,\chi)}(i,w) = \begin{cases} N(i,w) \cup \{||\chi||^M_{i,w}\} & \text{if } M,w \models \mathsf{B}_i\psi \wedge \\ & \mathsf{K}_i(\psi \to \chi) \\ N(i,w) & \text{otherwise} \end{cases}$$

$$N^{\cap(\psi,\chi)}(i,w) = \begin{cases} N(i,w) \cup \{||\psi \wedge \chi||^M_{i,w}\} & \text{if } M,w \models \mathsf{B}_i\psi \wedge \\ & \mathsf{B}_i\chi \\ N(i,w) & \text{otherwise} \end{cases}$$

Note that in the mono-agent case, thanks to Constraint (**C1**), we can assume, in a given model $M = (W, N, R_1, V)$, that R_1 is the universal relation on W.

According to the previous truth conditions, an agent i explicitly believes φ at world w if and only if, at world w, agent i has the fact corresponding to the formula φ (i.e., $||\varphi||^M_{i,w}$) included in her explicit belief set. Moreover, an agent has background knowledge that φ is true if and only if φ is true in all situations that are included in the agent's epistemic state. Mental operations of the form α are formalized as model update operations that expand or contract the agents' explicit belief sets. In particular, the mental operation $+\psi$ consists in perceiving ψ and adding it to the explicit belief set, while the mental operation $-\psi$ consists in forgetting ψ and removing it from the explicit belief set. The mental operation $\vdash(\psi,\chi)$ consists in adding the explicit belief χ to an agent's explicit belief set if the agent believes ψ and has background knowledge that ψ implies χ. The mental operation $\cap(\psi,\chi)$ consists in adding the explicit belief $\psi \wedge \chi$ to an agent's explicit belief set if the agent explicitly believes both ψ and χ.

We write $\models_{\mathsf{DLEK}} \varphi$ to denote that φ is valid, i.e., φ is true at every world w of every LEK-model M. In the next section we show some interesting validities of the logic DLEK.

3.3 Some validities

The following four validities capture the basic properties of the four mental operations $\vdash(\varphi,\psi)$, $\cap(\varphi,\psi)$, $+\varphi$ and $-\varphi$ semantically defined above. Let $\varphi, \psi \in Prop$. Then:

$$\models_{\mathsf{DLEK}} (\mathsf{K}_i(\varphi \to \psi) \wedge \mathsf{B}_i\varphi) \to [\vdash(\varphi,\psi)]\mathsf{B}_i\psi \quad (1)$$

$$\models_{\mathsf{DLEK}} (\mathsf{B}_i\varphi \wedge \mathsf{B}_i\psi) \to [\cap(\varphi,\psi)]\mathsf{B}_i(\varphi \wedge \psi) \quad (2)$$

$$\models_{\mathsf{DLEK}} [+\varphi]\mathsf{B}_i\varphi \quad (3)$$

$$\models_{\mathsf{DLEK}} [-\varphi]\neg\mathsf{B}_i\varphi \quad (4)$$

For instance, according to the first validity, if φ and ψ are propositional formulas, agent i explicitly believes φ and has background knowledge that φ implies ψ then, as a consequence of the mental operation $\vdash(\varphi,\psi)$, she will start to believe ψ. According to the third validity, if φ is a propositional formula then, as a consequence of perceiving that φ is true, agent i starts to explicitly believe that φ is true.

The reason why we need to impose that φ and ψ are propositional formulas is that there are DLEK-formulas such as the Moore-like formula $p \wedge \neg\mathsf{B}_i p$ for which the previous four principles do not hold. For instance, the following formula is not valid:

$$[+(p \wedge \neg\mathsf{B}_i p)]\mathsf{B}_i(p \wedge \neg\mathsf{B}_i p)$$

It is worth noting that in the logic DLEK we can 'simulate' in a dynamic way the rule of necessitation. Indeed:

$$\models_{\mathsf{DLEK}}\varphi \text{ implies } \models_{\mathsf{DLEK}} [+\top]\mathsf{B}_i\varphi \quad (5)$$

We can also dynamically 'simulate' Axiom K for the explicit belief operator, in the case of propositional formulas. Let $\varphi, \psi \in Prop$. Then, we have the following validity:

$$\models_{\mathsf{DLEK}}(\mathsf{B}_i\varphi \wedge \mathsf{B}_i(\varphi \to \psi)) \to [\cap(\varphi,\varphi \to \psi)][\vdash(\varphi \wedge \psi,\psi)]\mathsf{B}_i\psi \quad (6)$$

4. EXAMPLE

In this section we are going to illustrate our logic DLEK with the help of a concrete example. We are interested in describing the dynamics of explicit beliefs in a multi-agent scenario.

The scenario goes as follows. There are two robotic assistants, say robot A (Anne) and robot B (Bob), who have to take care of a person. The two robots are resource-bounded in the sense that they have background knowledge and (non-omniscient) explicit beliefs and form new explicit beliefs by means of the mental operations highlighted in Figure 1 in Section 2. The person communicates with the robots via a coloured electric light which can be either red or green. The communication code is the following one: (i) if the electric light is red (atom r) then it means that the person needs help (atom h), and (ii) if the electric is green (atom g) then it means that the person is having a rest and wants not to be bothered (atom b). We assume that:

H1. robot A has full knowledge about the communication code as she knows that r implies h and that g implies b,

H2. robot B has only partial knowledge about the communication code as he knows that r implies h but he does not know that g implies b, and

H3. robot A and robot B have common knowledge that they share a part of the communication code, namely that each of them knows that r implies h.

Thus, let us suppose that $Agt = \{A, B\}$ and $Atm = \{r, g, h, b\}$. We represent the initial situation by the minimal LEK model satisfying the previous hypothesis H1, H2 and H3 and which only excludes the impossible situations in which the electric light is both red and green and the person needs help and takes a rest at the same time. This model is the tuple $\mathtt{MR} = (W, N, R_A, R_B, V)$ (where $\mathtt{MR}$ stands for 'model of the robots') such that:

- $W = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$;
- $N(i,w) = \emptyset$ for all $i \in Agt$ and for all $w \in W$;
- the quotient set of W by R_A is $\{\{w_1, w_2, w_3, w_4, w_5\}, \{w_6, w_7\}, \{w_8, w_9\}\}$;

- the quotient set of W by R_B is $\{\{w_1, w_2, w_3, w_4, w_5, w_6, w_7\}, \{w_8, w_9\}\}$;

- $V(w_1) = \{r, h\}$ $V(w_2) = \{h\}$ $V(w_3) = \{g, b\}$
  $V(w_4) = \{b\}$ $V(w_5) = \emptyset$ $V(w_6) = \{g\}$
  $V(w_7) = \{g, h\}$ $V(w_8) = \{r\}$ $V(w_9) = \{r, b\}$.

The fact that $N(i, w) = \emptyset$ for all $i \in Agt$ and for all $w \in W$ just means that in the initial situation the robots do not have any explicit belief in their short-term memories.

Let us assume that w_1 is the actual situation in which the electric light is red and the person needs help.

The first thing to observe is that in the actual situation the hypothesis H1 and H2 are both satisfied. Indeed, as for H1, we have:

$$\mathtt{MR}, w_1 \models \mathsf{K}_A(r \to h) \wedge \mathsf{K}_A(g \to b).$$

As for H2, we have:

$$\mathtt{MR}, w_1 \models \mathsf{K}_B(r \to h) \wedge \neg\mathsf{K}_B(g \to b).$$

In order to specify hypothesis H3, let $\mathsf{EK}_J\varphi$ be an abbreviation of $\bigwedge_{i \in J} \mathsf{K}_i\varphi$, standing for "every agent in J has background knowledge that φ". Then, let us define $\mathsf{EK}^k_J\varphi$ by induction for every natural number $k \in \mathbb{N}$:

$$\mathsf{EK}^0_J\varphi \stackrel{\text{def}}{=} \varphi$$

and for all $k \geq 1$:

$$\mathsf{EK}^k_J\varphi \stackrel{\text{def}}{=} \mathsf{EK}_J(\mathsf{EK}^{k-1}_J\varphi).$$

For every natural number $n \in \mathbb{N}$, let $\mathsf{MK}^n_J\varphi$ be an abbreviation of $\bigwedge_{1 \leq k \leq n} \mathsf{EK}^k_J\varphi$. $\mathsf{MK}^n_J\varphi$ expresses J's background mutual knowledge that φ up to n iterations, *i.e.*, everyone in J has background knowledge that φ, everyone in J has background knowledge that everyone in J has background knowledge that φ, and so on until level n. We omit the full definitions of the universal explicit belief operator EB_J (with $\mathsf{EB}_J\varphi$ standing for "every agent in J explicitly believes that φ") and of the mutual explicit belief operator MB^n_J (with $\mathsf{MB}^n_J\varphi$ standing for "the agents in J have mutual explicit belief that φ up to n iterations") since they can be defined exactly in the same way as the universal knowledge and mutual knowledge operators, starting from the individual explicit knowledge operators.

The previous hypothesis H3 holds in the initial situation since, for every $n \in \mathbb{N}$, we have:

$$\mathtt{MR}, w_1 \models \mathsf{MK}^n_{\{A,B\}}\mathsf{EK}_{\{A,B\}}(r \to h).$$

Let us suppose that the person switches on the red light in order to signal to the two robots that she needs help. This event is represented by the mental operation $+r$ which leads from model $\mathtt{MR}$ to the updated model $\mathtt{MR}^{+r} = (W, N^{+r}, R_A, R_B, V)$ such that:

$N^{+r}(A, w_1) = \{\{w_1\}\}$, $N^{+r}(A, w_2) = \{\{w_1\}\}$,
$N^{+r}(A, w_3) = \{\{w_1\}\}$, $N^{+r}(A, w_4) = \{\{w_1\}\}$,
$N^{+r}(A, w_5) = \{\{w_1\}\}$, $N^{+r}(A, w_6) = \emptyset$,
$N^{+r}(A, w_7) = \emptyset$, $N^{+r}(A, w_8) = \{\{w_8, w_9\}\}$,
$N^{+r}(A, w_9) = \{\{w_8, w_9\}\}$, $N^{+r}(B, w_1) = \{\{w_1\}\}$,
$N^{+r}(B, w_2) = \{\{w_1\}\}$, $N^{+r}(B, w_3) = \{\{w_1\}\}$,
$N^{+r}(B, w_4) = \{\{w_1\}\}$, $N^{+r}(B, w_5) = \{\{w_1\}\}$,
$N^{+r}(B, w_6) = \{\{w_1\}\}$, $N^{+r}(B, w_7) = \{\{w_1\}\}$,
$N^{+r}(B, w_8) = \{\{w_8, w_9\}\}$, $N^{+r}(B, w_9) = \{\{w_8, w_9\}\}$.

It is easy to check that in the new situation, after the mental operation $+r$ has been executed, the two robots explicitly believe that the light is red. That is:

$$\mathtt{MR}^{+r}, w_1 \models \mathsf{EB}_{\{A,B\}}r.$$

However, the mental operation is not sufficient to guarantee that the robots believe that the person needs helps. Indeed, we have:

$$\mathtt{MR}^{+r}, w_1 \models \neg\mathsf{B}_A h \wedge \neg\mathsf{B}_B h.$$

It is by trying to infer that the person needs help from the fact that she switched on the red light, represented by $\vdash(r,h)$, that the robots can form this explicit belief. The mental operation $\vdash(r,h)$ leads from model $\mathtt{MR}^{+r}$ to the updated model $(\mathtt{MR}^{+r})^{\vdash(r,h)} = (W, (N^{+r})^{\vdash(r,h)}, R_A, R_B, V)$ such that:

$(N^{+r})^{\vdash(r,h)}(A, w_1) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_2) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_3) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_4) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_5) = \{\{w_1\}, \{w_1, w_2\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_6) = \{\emptyset\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_7) = \{\emptyset\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_8) = \{\{w_8, w_9\}\}$,
$(N^{+r})^{\vdash(r,h)}(A, w_9) = \{\{w_8, w_9\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_1) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_2) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_3) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_4) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_5) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_6) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_7) = \{\{w_1\}, \{w_1, w_2, w_7\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_8) = \{\{w_8, w_9\}\}$,
$(N^{+r})^{\vdash(r,h)}(B, w_9) = \{\{w_8, w_9\}\}$.

It is easy to check that in the new situation, after the mental operation $\vdash(r,h)$ has been executed, the two robots explicitly believe that the person needs help. That is:

$$(\mathtt{MR}^{+r})^{\vdash(r,h)}, w_1 \models \mathsf{EB}_{\{A,B\}}h.$$

To sum up, we have the following which holds:

$$\mathtt{MR}, w_1 \models [+r]\mathsf{EB}_{\{A,B\}}r \wedge [+r](\neg\mathsf{B}_A h \wedge \neg\mathsf{B}_B h) \wedge [+r][\vdash(r,h)]\mathsf{EB}_{\{A,B\}}h.$$

It is worth noting that the sequence of the two mental operations $+r$ and $\vdash(r,h)$ is not sufficient for the robots to acquire the explicit common belief that the person needs help. Indeed, it is to check that the following holds:

$$(\mathtt{MR}^{+r})^{\vdash(r,h)}, w_1 \models \neg\mathsf{B}_A\mathsf{B}_B h \wedge \neg\mathsf{B}_B\mathsf{B}_A h.$$

In order to obtain such explicit common belief, the two robots need to perform the two additional mental operations $\vdash(h,\mathsf{B}_A h)$ and $\vdash(h,\mathsf{B}_B h)$ (i.e., trying to infer that robot A/robot B explicitly believes that h from the fact that h holds), no matter the order in which they are executed. Indeed, for every $n \in \mathbb{N}$, we have:

$$\mathtt{MR}, w_1 \models [+r][\vdash(r,h)][\vdash(r,\mathsf{B}_A h)][\vdash(r,\mathsf{B}_B h)]\mathsf{MB}^n_{\{A,B\}}h,$$
$$\mathtt{MR}, w_1 \models [+r][\vdash(r,h)][\vdash(r,\mathsf{B}_B h)][\vdash(r,\mathsf{B}_A h)]\mathsf{MB}^n_{\{A,B\}}h.$$

It is just routine to check that the mental operations $+g$ and $\vdash(g,b)$ are also sufficient for robot A to start to believe b explicitly after performing them, but they are not sufficient

for robot B since he does not have background knowledge that g implies b. In formal terms, we have:

$$\mathtt{MR}, w_1 \models [+g][\vdash(g,b)]\mathsf{B}_A b,$$
$$\mathtt{MR}, w_1 \models [+g][\vdash(g,b)]\neg\mathsf{B}_B b.$$

5. AXIOMATIZATION

Let us now present sound and complete axiomatizations for the logic LEK and its dynamic extension DLEK.

DEFINITION 3. *We define LEK to be the extension of classical propositional logic given by the following rules and axioms:*

$$(\mathsf{K}_i\varphi \wedge \mathsf{K}_i(\varphi \to \psi)) \to \mathsf{K}_i\psi \qquad (\mathbf{K}_{\mathsf{K}_i})$$
$$\mathsf{K}_i\varphi \to \varphi \qquad (\mathbf{T}_{\mathsf{K}_i})$$
$$\mathsf{K}_i\varphi \to \mathsf{K}_i\mathsf{K}_i\varphi \qquad (\mathbf{4}_{\mathsf{K}_i})$$
$$\neg\mathsf{K}_i\varphi \to \mathsf{K}_i\neg\mathsf{K}_i\varphi \qquad (\mathbf{5}_{\mathsf{K}_i})$$
$$(\mathsf{B}_i\varphi \wedge \mathsf{K}_i(\varphi \leftrightarrow \psi)) \to \mathsf{B}_i\psi \qquad (\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$$
$$\mathsf{B}_i\varphi \to \mathsf{K}_i\mathsf{B}_i\varphi \qquad (\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$$
$$\frac{\varphi}{\mathsf{K}_i\varphi} \qquad (\mathbf{Nec}_{\mathsf{K}_i})$$

The axiomatics of the logic DLEK includes all principles of the logic LEK *plus* a set of reduction axioms and the rule of replacement of equivalents.

DEFINITION 4. *We define DLEK to be the extension of LEK generated by the following axioms:*

$$[\alpha]p \leftrightarrow p \qquad (\mathbf{Red}_p)$$
$$[\alpha]\neg\varphi \leftrightarrow \neg[\alpha]\varphi \qquad (\mathbf{Red}_\neg)$$
$$[\alpha](\varphi \wedge \psi) \leftrightarrow ([\alpha]\varphi \wedge [\alpha]\psi) \qquad (\mathbf{Red}_\wedge)$$
$$[\alpha]\mathsf{K}_i\varphi \leftrightarrow \mathsf{K}_i[\alpha]\varphi \qquad (\mathbf{Red}_{\mathsf{K}_i})$$
$$[+\varphi]\mathsf{B}_i\psi \leftrightarrow (\mathsf{B}_i[+\varphi]\psi \vee \mathsf{K}_i([+\varphi]\psi \leftrightarrow \varphi)) \qquad (\mathbf{Red}_{\mathsf{B}_i,+})$$
$$[-\varphi]\mathsf{B}_i\psi \leftrightarrow (\mathsf{B}_i[-\varphi]\psi \vee \neg\mathsf{K}_i([-\varphi]\psi \leftrightarrow \varphi)) \qquad (\mathbf{Red}_{\mathsf{B}_i,-})$$
$$[\vdash(\varphi,\psi)]\mathsf{B}_i\chi \leftrightarrow \Big(\mathsf{B}_i[\vdash(\varphi,\psi)]\chi \vee \big((\mathsf{B}_i\varphi \wedge \mathsf{K}_i(\varphi \to \psi)) \wedge \mathsf{K}_i([\vdash(\varphi,\psi)]\chi \leftrightarrow \psi)\big)\Big) \qquad (\mathbf{Red}_{\mathsf{B}_i,\vdash})$$
$$[\cap(\varphi,\psi)]\mathsf{B}_i\chi \leftrightarrow \Big(\mathsf{B}_i[\cap(\varphi,\psi)]\chi \vee \big((\mathsf{B}_i\varphi \wedge \mathsf{B}_i\psi) \wedge \mathsf{K}_i([\cap(\varphi,\psi)]\chi \leftrightarrow (\varphi \wedge \psi))\big)\Big) \qquad (\mathbf{Red}_{\mathsf{B}_i,\cap})$$

and the following rule of inference:

$$\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]} \qquad (\mathbf{RRE})$$

We write $\mathsf{DLEK} \vdash \varphi$ to denote the fact that φ is a theorem of DLEK.

It is straightforward to check that all axioms are valid and all rules preserve validity in the class of DLEK-models, from which the following is an immediate consequence:

LEMMA 1. *The logics LEK and DLEK are sound for the class of LEK-models.*

Our goal now is to prove that LEK is strongly complete for its intended semantics. We will achieve this by a fairly standard canonical-model argument, although the neighborhood structure will require some care.

DEFINITION 5 (CANONICAL LEK MODEL). *We define the canonical LEK model*

$$M_c = (W_c, N_c, R_{c1} \ldots, R_{cn}, V_c)$$

where:

- W_c *is the set of all maximal consistent subsets of* $\mathcal{L}_{\mathsf{LEK}}$.
- $wR_{ci}v$ *if and only if, for all formulas* φ *and all agents* i, $\mathsf{K}_i\varphi \in w$ *if and only if* $\mathsf{K}_i\varphi \in v$.
- *In order to define* N_c, *for* $w \in W$ *and* $\varphi \in \mathcal{L}_{\mathsf{LEK}}$, *first define*

$$A_\varphi(i,w) = \{v \in R_i(w) : \varphi \in v\}.$$

Then, define N_c *by letting*

$$N_c(i,w) = \{A_\varphi(i,w) : \mathsf{B}_i\varphi \in w\}.$$

- *Finally, we define the valuation* V_c *by* $w \in V_c(p)$ *if and only if* $p \in w$.

The following is standard and we omit the proof:

LEMMA 2. *The structure* M_c *defined above is a* $\mathcal{L}_{\mathsf{LEK}}$*-model. Moreover, if* $w \in W_c$ *and* $\varphi \in \mathcal{L}_{\mathsf{LEK}}$, *then*

1. $\mathsf{K}_i\varphi \in w$ *if and only if, for every* v *such that* $wR_{ci}v$, $\varphi \in v$, *and*
2. *if* $wR_{ci}v$ *and* $\mathsf{B}_i\varphi \in w$, *then* $\mathsf{B}_i\varphi \in v$.

We also need to prove that M_c has a somewhat less familiar property. This will be used later in the truth lemma, for the case of B_i.

LEMMA 3. *For every* $w \in W_c$ *and* $\mathsf{B}_i\varphi, \mathsf{B}_i\psi \in \mathcal{L}_{\mathsf{LEK}}$, *if* $\mathsf{B}_i\varphi \in w$ *but* $\mathsf{B}_i\psi \notin w$, *it follows that there is* $v \in R_{ci}(w)$ *such that either* $\varphi \in v$ *but* $\neg\psi \in v$, *or* $\neg\varphi \in v$ *but* $\psi \in v$.

PROOF. Let $w \in W_c$ and φ, ψ be such that $\mathsf{B}_i\varphi \in w$, $\mathsf{B}_i\psi \notin w$. Towards a contradiction, assume that for every $v \in R_{ci}(w)$, either $\varphi, \psi \in v$ or $\neg\varphi, \neg\psi \in v$; then, it follows from Lemma 2 that $\mathsf{K}_i(\varphi \leftrightarrow \psi) \in w$, so that by Axiom $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$, $\mathsf{B}_i\psi \in w$, contrary to our assumption. □

With this, we may state and prove our version of the Truth Lemma:

LEMMA 4. *For every* $\varphi \in \mathcal{L}_{\mathsf{LEK}}$ *and every* $w \in W_c$, $\varphi \in w$ *if and only if* $M_c, w \models \varphi$.

PROOF. The proof proceeds by a standard induction on the construction of φ. All cases are routine except $\varphi = \mathsf{B}_i\psi$.

First assume that $\mathsf{B}_i\psi \in w$. Then, $A_\psi(i,w) \in N_c(i,w)$. But,

$$A_\psi(i,w) = \{v \in R_i(w) : \psi \in v\} \overset{\text{IH}}{=} \|\psi\| \cap R_i(w).$$

Thus, $M_c, w \models \mathsf{B}_i\psi$.

Now, suppose $\mathsf{B}_i\psi \notin w$, so that $\neg\mathsf{B}_i\psi \in w$. We must check that $\|\psi\| \cap R_{ci}(w) \notin N(i,w)$. Choose an arbitrary set $A \in N(i,w)$; by definition, $A = A_\theta(i,w)$ for some θ with

$\mathsf{B}_i\theta \in w$. By Lemma 3, there is some $v \in R_i(w)$ such that $\psi, \neg\theta \in v$ or $\neg\psi, \theta \in v$; in the first case, this shows using the induction hypothesis that $v \in (\|\psi\| \cap R_i(w)) \setminus A_\theta(i,w)$, while in the second we obtain $v \in A_\theta(i,w) \setminus (\|\psi\| \cap R_i(w))$. In either case we obtain $A_\theta(i,w) \neq \|\psi\| \cap R_i(w)$, and since $A = A_\theta(i,w)$ was an arbitrary element of $N(i,w)$, we conclude that $\|\psi\| \cap R_i(w) \notin N(i,w)$ and thus $M_c, w \not\models \mathsf{B}_i\psi$. □

We are now ready to prove that the static LEK is strongly complete.

THEOREM 1. *LEK is strongly complete for the class of LEK models.*

PROOF. Any consistent set of formulas Φ may be extended to a maximal consistent set of formulas $w_* \in W_c$, and $M_c, w_* \models \Phi$ by Lemma 4. □

The strong completeness of DLEK follows from this result, in view of the fact that the reduction axioms may be used to find, for any DLEK formula, a provably equivalent LEK formula.

LEMMA 5. *If φ is any formula of $\mathcal{L}_{DLEK}$, there is a formula $\widetilde{\varphi}$ in $\mathcal{L}_{LEK}$ such that DLEK $\vdash \varphi \leftrightarrow \widetilde{\varphi}$.*

PROOF. This follows by a routine induction on φ using the reduction axioms and the rule of replacement of equivalents (**RRE**) from Definition 4. □

As a corollary, we get the following:

THEOREM 2. *DLEK is strongly complete for the class of LEK models.*

PROOF. If Γ is a consistent set of $\mathcal{L}_{\mathsf{DLEK}}$ formulas, then $\widetilde{\Gamma} = \{\widetilde{\varphi} : \varphi \in \Gamma\}$ is a consistent set of $\mathcal{L}_{\mathsf{LEK}}$ formulas (since DLEK is an extension of LEK), and hence by Theorem 1, there is a model M with a world w such that $M, w \models \widetilde{\Gamma}$. But, since DLEK is sound and for each $\varphi \in \Gamma$, DLEK $\vdash \varphi \leftrightarrow \widetilde{\varphi}$, it follows that $M, w \models \Gamma$. □

Thus our logics are strongly complete, but the construction we have given will in general produce infinite models. In the next section, we are going move from axiomatics to complexity of the satisfiability problem.

6. COMPLEXITY

In this section we will study the computability of the satisfiability problem of LEK: given a formula φ, determine whether φ is satisfiable. We will also provide a decidability result of the satisfiability problem of DLEK. We first consider the single-agent case and move then to the multi-agent case. Below, $card(A)$ denotes the cardinality of the set A.

6.1 Mono-agent case

Firstly, assume $card(Agt) = 1$. Let φ be a satisfiable formula. Let $M = (W, N, V)$ be a model and $w \in W$ be such that $M, w \models \varphi$. Let $\mathsf{K}\psi_1, \ldots, \mathsf{K}\psi_m$ and $\mathsf{B}\chi_1, \ldots, \mathsf{B}\chi_n$ be lists of the set of all subformulas of φ of the form $\mathsf{K}\psi$ and $\mathsf{B}\chi$. Let $\hat{\mathsf{K}} = \{i : 1 \leq i \leq m \ \&\ M, w \not\models \mathsf{K}\psi_i\}$, $\hat{\mathsf{B}}^+ = \{j : 1 \leq j \leq n \ \&\ M, w \models \mathsf{B}\chi_j\}$ and $\hat{\mathsf{B}}^- = \{k : 1 \leq k \leq n \ \&\ M, w \not\models \mathsf{B}\chi_k\}$. For all $i \in \hat{\mathsf{K}}$, let $v_i \in W$ be such that $M, v_i \not\models \psi_i$. Such v_i exists because $M, w \not\models \mathsf{K}\psi_i$. For all $j \in \hat{\mathsf{B}}^+$, let $A_j \in N$ be such that $A_j = \{v \in W : M, v \models \chi_j\}$. Such A_j exists because $M, w \models \mathsf{B}\chi_j$. For all $j \in \hat{\mathsf{B}}^+$ and for all $k \in \hat{\mathsf{B}}^-$, let $u_{j,k} \in W$ be such that $M, u_{j,k} \not\models \chi_j \leftrightarrow \chi_k$. Such $u_{j,k}$ exists because $M, w \models \mathsf{B}\chi_j$ and $M, w \not\models \mathsf{B}\chi_k$. Let $M' = (W', N', V')$ be the model defined as follows:

- $W' = \{w\} \cup \{v_i : i \in \hat{\mathsf{K}}\} \cup \{u_{j,k} : j \in \hat{\mathsf{B}}^+ \ \&\ k \in \hat{\mathsf{B}}^-\}$,
- $N' = \{A_j \cap W' : j \in \hat{\mathsf{B}}^+\}$,
- for all $p \in VAR$, $V'(p) = V(p) \cap W'$.

Obviously, $card(W')$ and $card(N')$ are polynomial in the size of φ. Let Φ be the closure under single negations of the set of all φ's subformulas.

LEMMA 6. *Let φ be a formula. If $\varphi \in \Phi$ then for all $s \in W'$, $M, s \models \varphi$ iff $M', s \models \varphi$.*

PROOF. By induction on φ. We only consider the cases $\varphi = \mathsf{B}\chi$.
Suppose $M, s \models \mathsf{B}\chi$ and $M', s \not\models \mathsf{B}\chi$. Let $j \in \hat{\mathsf{B}}^+$ be such that $\mathsf{B}\chi = \mathsf{B}\chi_j$. Hence, $A_j = \{t \in W : M, t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in W' : M', t \models \chi\}$. Thus, $M', s \models \mathsf{B}\chi$: a contradiction.
Suppose $M', s \models \mathsf{B}\chi$ and $M, s \not\models \mathsf{B}\chi$. Let $j \in \hat{\mathsf{B}}^+$ be such that $A_j \cap W' = \{t \in W' : M', t \models \chi\}$. By induction hypothesis, $A_j \cap W' = \{t \in W' : M, t \models \chi\}$. Let $k \in \hat{\mathsf{B}}^-$ be such that $\mathsf{B}\chi = \mathsf{B}\chi_k$. Remember that $A_j = \{t \in W : M, t \models \chi_j\}$. Moreover, $u_{j,k} \in W'$ is such that $M, u_{j,k} \not\models \chi_j$ and $M, u_{j,k} \not\models \chi_k$, or $M, u_{j,k} \not\models \chi_j$ and $M, u_{j,k} \models \chi_k$. In the former case, $u_{j,k} \in A_j \cap W'$. Hence, $M, u_{j,k} \models \chi$: a contradiction. In the latter case, $u_{j,k} \in A_j \cap W'$. Thus, $M, u_{j,k} \models \chi_j$: a contradiction. □

THEOREM 3. *If $card(Agt) = 1$ then satisfiability problem of LEK is NP-complete.*

PROOF. Membership in NP follows from Lemma 6. NP-hardness follows from the NP-hardness of classical propositional logic. □

6.2 Multi-agent case

Our study of the computability in the multi-agent case will be based on the modal tableaux approach developed by Halpern and Moses [11]. Assume $card(Agt) \geq 2$. For all formulas φ, let $SF(\varphi)$ be the closure under single negations of the set of all subformulas of φ.

Modal tableaux.

A set $\mathcal{T}$ of formulas is said to be fully expanded if for all formulas φ, if $\varphi \in \mathcal{T}$ then for all formulas ψ, if $\psi \in SF(\varphi)$ then either $\psi \in \mathcal{T}$, or $\neg\psi \in \mathcal{T}$, $SF(\varphi)$ denoting the closure under single negations of the set of all φ's subformulas. A propositional tableau is a set $\mathcal{T}$ of formulas such that the following conditions holds: (i) for all formulas φ, if $\neg\neg\varphi \in \mathcal{T}$ then $\varphi \in \mathcal{T}$; (ii) for all formulas φ, ψ, if $\neg(\varphi \vee \psi) \in \mathcal{T}$ then $\neg\varphi \in \mathcal{T}$ and $\neg\psi \in \mathcal{T}$; (iii) for all formulas φ, ψ, if $\varphi \vee \psi \in \mathcal{T}$ then either $\varphi \in \mathcal{T}$, or $\psi \in \mathcal{T}$. A propositional tableau $\mathcal{T}$ is said to be blatantly consistent iff for all formulas φ, either $\varphi \notin \mathcal{T}$, or $\neg\varphi \notin \mathcal{T}$.

A modal tableau is a structure of the form $\mathcal{T} = (W, R, L)$ where W is a nonempty set of states (with typical members denoted w, v, etc), R is a function associating a binary relation R_i on W to each $i \in Agt$ and L is a function assigning to each $w \in W$ a blatantly consistent and fully expanded propositional tableau $L(w)$ such that for all $w \in W$, the following conditions holds: (i) for all formulas φ, if $\neg \mathsf{K}_i\varphi \in L(w)$ then there exists $v \in R_i(w)$ such that $\neg\varphi \in L(v)$; (ii) for all formulas φ, if $\mathsf{B}_i\varphi \in L(w)$ then for all formulas ψ, if $\neg \mathsf{B}_i\psi \in L(w)$ then there exists $v \in R_i(w)$ such that either $\varphi \in L(v)$ and $\neg\psi \in L(v)$, or $\neg\varphi \in L(v)$ and $\psi \in L(v)$; (iii) for all formulas φ, if $\mathsf{K}_i\varphi \in L(w)$ then for all $v \in (R_i \cup R_i^{-1})^\star(w)$, $\varphi \in L(v)$ and $\mathsf{K}_i\varphi \in L(v)$; (iv) for all formulas φ, if $\mathsf{B}_i\varphi \in L(w)$ then for all $v \in (R_i \cup R_i^{-1})^\star(w)$, $\mathsf{B}_i\varphi \in L(v)$. For all formulas φ, let $L^{-1}(\varphi) = \{w : \ w \in W \ \& \ \varphi \in L(w)\}$. We shall say that a modal tableau $\mathcal{T} = (W, R, L)$ is a modal tableau for a formula φ if $L^{-1}(\varphi) \neq \emptyset$.

From models to modal tableaux.

Given a model $M = (W, R, N, V)$, let $\mathcal{T}' = (W', R', L')$ be defined as follows: $W' = W$, $R'_i = R_i$ for each $i \in Agt$, L' is the function assigning to each $w \in W'$ the propositional tableau $L'(w) = \{\varphi : \ M, w \models \varphi\}$. The proof that $\mathcal{T}'$ is a modal tableau is easy. As a result,

PROPOSITION 1. *Let φ be a formula. If φ is satisfiable then there exists a modal tableau for φ.*

From modal tableaux to models.

Given a modal tableau $\mathcal{T} = (W, R, L)$, let

$$M' = (W', R', N', V')$$

be defined as follows: $W' = W$, $R'_i = (R_i \cup R_i^{-1})^\star$ for each $i \in Agt$, $N'_i(w) = \{(R_i \cup R_i^{-1})^\star(w) \cap L^{-1}(\varphi) : \ \mathsf{B}_i\varphi \in L(w)\}$ for each $w \in W$ and for each $i \in Agt$, V' is the function assigning to each $p \in Atm$ the subset $V'(p) = \{w : \ w \in W \ \& \ p \in L(w)\}$ of W'. The proof that $\mathcal{T}'$ is a model is easy. Moreover, one can prove by induction on φ that for all $w \in W$, $\varphi \in L(s)$ iff $M', s \models \varphi$. As a result,

PROPOSITION 2. *Let φ be a formula. If there exists a modal tableau for φ then φ is satisfiable.*

Membership in PSPACE.

By Propositions 1 and 2, satisfiability is reducible to the following decision problem (MT): given a formula φ, determine whether there exists a modal tableau for φ. Based on the tools and techniques developed in ordinary epistemic logics by Halpern and Moses [11], one can design an algorithm that tries to construct a modal tableau for a given formula φ. The main properties of such algorithm are:

- For all given formulas φ, the above algorithm terminates and runs in polynomial space,
- for all given formulas φ, the algorithm returns "there is a modal tableau for φ" iff there is a modal tableau for φ.

PROPOSITION 3. *There is an algorithm for deciding satisfiability that runs in polynomial space.*

As a result,

THEOREM 4. *If $card(Agt) \geq 2$ then satisfiability problem of LEK is PSPACE-complete.*

PROOF. Membership in PSPACE follows from the above discussion. PSPACE-hardness follows from the PSPACE-hardness of multi-agent epistemic logic. □

The reduction axioms and the rule of replacement of equivalents in Definition 4 may then be used to give a decision procedure for the satisfiability of DLEK; however, due to exponential blow-up in the size of formulas, this algorithm would no longer remain in PSPACE without modification. Thus we will state only the following:

COROLLARY 1. *The satisfiability problem of DLEK is decidable.*

PROOF. Immediate from the decidability of LEK and the fact that, given a formula φ of $\mathcal{L}_{\mathsf{DLEK}}$, the formula $\widetilde{\varphi}$ is clearly computable from φ. □

However, we do not believe that such a procedure would be optimal, and indeed conjecture that DLEK is in PSPACE. We leave the computation of its precise complexity for future work.

7. CONCLUSION

Let's take stock. In the paper we have introduced DLEK, a logical theory of belief dynamics for resource-bounded agents inspired by existing psychological theories of human memory. We have provided decidability and complexity results for DLEK as well as for its static fragment LEK.

Directions of future research are manifold. On the conceptual level, we plan to complete the conceptual framework described in Section 2 by extending the family of mental operations with operations of *storage* of information in long-term memory. We believe that these kinds of operations can be modelled as special kinds of epistemic actions in the DEL sense. Specifically, a storage operation modifies an agent's background knowledge by restricting the epistemic relation R_i to worlds in which the information φ to be stored is true, under the condition that this information is already available in the agent's working memory and explicitly believed by the agent.

On the technical level, as emphasized in Section 6, we plan to obtain a result about complexity of the satisfiability problem of DLEK. In particular, we plan to prove that this problem is PSPACE-complete by appropriately adapting to our framework the technique proposed by Lutz [18] for studying complexity of the satisfiability problem of public announcement logic (PAL). We also plan to refine our approach by relaxing the assumption that mental operations are performed by all agents and that this is a public fact. To this aim, we will have to introduce a notion of action model in this sense of [4] which allows us to model a *private* form of mental operation (i.e., an operation which occurs in the mind of a specific agent without the other agents being aware of this).

REFERENCES

[1] T. Ågotnes and N. Alechina. A logic for reasoning about knowledge of unawareness. *Journal of Logic, Language and Information*, 23(2):197–217, 2014.

[2] N. Alechina, B. Logan, and M. Whitsey. A complete and decidable logic for resource-bounded agents. In *Proceedings of AAMAS 2014*, pages 606–613. IEEE Computer Society, 2004.

[3] A. D. Baddeley and G. Hitch. Working memory. In G. H. Bower, editor, *The psychology of learning and motivation: Advances in research and theory*, pages 47–89. Academic Press, 1974.

[4] A. Baltag and L. S. Moss. Logics for Epistemic Programs. *Synthese*, 139(2):165–224, 2004.

[5] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In *Proceedings of LOFT 7*, pages 13–60. Amsterdam University Press, 2008.

[6] N Cowan. An embedded process model of working memory. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 62–101. Cambridge University Press, 1999.

[7] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[8] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[9] J. Grant, S. Kraus, and D. Perlis. A logic for characterizing multiple bounded agents. *Autonomous Agents and Multi-Agent Systems*, 3(4):351–387, 2000.

[10] D. Grossi and F. R. Velázquez-Quesada. Twelve Angry Men: A study on the fine-grain of announcements. In *Proceedings of LORI 2009*, volume 5834 of *LNCS*, pages 147–160. Springer, 2009.

[11] J. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[12] J. Y. Halpern and L. C. Rêgo. Reasoning about knowledge of unawareness. *Games and Economic Behavior*, 67(2):503–525, 2009.

[13] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.

[14] D. Kahneman and A. Tversky. Variants of uncertainty. *Cognition*, 11:143–157, 1982.

[15] J. E. Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.

[16] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of AAAI-84*, pages 198–202. AAAI Press, 1984.

[17] M. C. Lovett, L. M. Reder, and C. Lebiere. Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 135–182. Cambridge University Press, 1999.

[18] C. Lutz. Complexity and Succinctness of Public Announcement Logic. In *Proceedings of AAMAS 2006*, pages 137–144. ACM Press, 2006.

[19] H. van Ditmarsch and T. French. Semantics for knowledge and change of awareness. *Journal of Logic, Language and Information*, 23(2):169–195, 2014.

[20] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2008.

[21] F. R. Velázquez-Quesada. Explicit and implicit knowledge in neighbourhood models. In *Proceedings of LORI 2013*, volume 8196 of *LNCS*, pages 239–252. Springer, 2013.

[22] R. M. Young and R. L. Lewis. The Soar cognitive architecture and human working memory. In A. Miyake and P. Shah, editors, *Models of working memory: Mechanisms of Active Maintenance and Executive Control*, pages 135–182. Cambridge University Press, 1999.